BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

MOLECULAR AND CELLULAR BIOLOGY, Feb. 1992, p. 716–723
0270-7306/92/020716-08$02.00/0

Vol. 12, No. 2

Complex Recognition Site for the Group I Intron-Encoded Endonuclease I-*Sce*II

CATHERINE WERNETTE,[1] ROLAND SALDANHA,[2] DAVID SMITH,[3] DING MING,[3] PHILIP S. PERLMAN,[1] AND RONALD A. BUTOW[1]*

Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235[1]; Department of Molecular Genetics, The Ohio State University, Columbus, Ohio 43210[2]; and Department of Molecular Biology and Genetics, Wayne State University School of Medicine, Detroit, Michigan 48201[3]

Received 28 August 1991/Accepted 15 November 1991

We have characterized features of the site recognized by a double-stranded DNA endonuclease, I-*Sce*II, encoded by intron 4α of the yeast mitochondrial *COX1* gene. We determined the effects of 36 point mutations on the cleavage efficiency of natural and synthetic substrates containing the *Saccharomyces capensis* I-*Sce*II site. Most mutations of the 18-bp I-*Sce*II recognition site are tolerated by the enzyme, and those mutant sites are cleaved between 42 and 100% as well as the wild-type substrate is. Nine mutants blocked cleavage to ≤33% of the wild-type, whereas only three point mutations, G-4→C, G-12→T, and G-15→C, block cleavage completely. Competition experiments indicate that these three substrates are not cleaved, at least in part because of a marked reduction in the affinity of the enzyme for those mutant DNAs. About 90% of the DNAs derived from randomization of the nucleotide sequence of the 4-bp staggered I-*Sce*II cleavage site are not cleaved by the enzyme. I-*Sce*II cleaves cloned DNA derived from human chromosome 3 about once every 110 kbp. The I-*Sce*II recognition sites in four randomly chosen human DNA clones have 56 to 78% identity with the 18-bp site in yeast mitochondrial DNA; they are cleaved at least 50% as well as the wild-type mitochondrial substrate despite the presence of some substitutions that individually compromise cleavage of the mitochondrial substrate. Analysis of these data suggests that the effect of a given base substitution on I-*Sce*II cleavage may depend on the sequence at other positions.

Recent studies show that a number of group I introns are mobile genetic elements (20). These mobile introns are found in diverse genomes including fungal mitochondria (5, 11, 15, 28), chloroplasts (7, 13, 14), nuclei (17, 18), and bacteriophages (1, 2, 21). In each case the intron has a reading frame that encodes an endonuclease function needed for intron mobility. Available data indicate that the intron-encoded endonuclease initiates the conversion or homing event (6) by introducing a double-strand break in the DNA of the recipient, intronless allele; that break is then repaired by using the intron-containing allele as a template. This process of intron homing is consistent with the double-strand break–gap repair model of gene conversion proposed by Szostak et al. (26).

The mitochondrial genome of *Saccharomyces cerevisiae* harbors two well-characterized homing group I introns, each encoding a site-specific endonuclease: the 1.1-kb ω intron of the 21S rRNA gene (3, 11, 15, 30) and al4α, the fourth intron of the *COX1* gene (5, 28). The ω-encoded endonuclease (I-*Sce*I) is the product of a free-standing open reading frame within the intron (3), whereas the al4α-encoded endonuclease (I-*Sce*II) is derived by proteolytic processing of a fusion polypeptide encoded by exons 1 to 4 of the *COX1* gene and the al4α open reading frame (9, 23, 29). Active I-*Sce*II endonuclease is composed largely, if not entirely, of amino acids encoded by al4α (5, 29). Although the substrates for these enzymes are unique, each makes a staggered double-strand break within a 4-base sequence (hereafter referred to as the cleavage site) leaving 3′-hydroxyl overhangs. I-*Sce*I cleaves symmetrically about the intron insertion site (4), whereas I-*Sce*II cleaves both strands to one side of the intron insertion site (5, 28). The overall sequence recognized by each of these enzymes includes an 18- to 20-bp sequence (hereafter referred to as the recognition site), situated asymmetrically about the site of intron insertion.

Universal code equivalent genes for both I-*Sce*I and I-*Sce*II have been expressed in *Escherichia coli*, and the endonuclease products of these genes have been partially characterized (3, 5, 16). For both enzyme activities, point mutations within their recognition sites result in a range of effects, from mutant substrates that are not cleaved at all to those whose cleavage is indistinguishable from that of the wild type. Sargueil et al. (22) also used I-*Sce*II activity present in crude mitochondrial extracts of wild-type yeast cells with the same set of mutant substrates and found that the recognition sites differed significantly from those defined with the activities obtained from *E. coli*. Moreover, these authors noted significant differences in cleavage specificity among three related but different forms of I-*Sce*II expressed in *E. coli* and the enzyme extracted from mitochondria of wild-type yeast cells. In those studies the substrate for I-*Sce*II was a cleavage site representing a region of fused exons 4 and 5 of the *COX1* gene of *S. cerevisiae*.

We recently described the purification of I-*Sce*II from mitochondria of a mutant yeast strain that overproduces the enzyme (29). We found that the active form of I-*Sce*II fractionates as a dimer of 31-kDa subunits encoded probably entirely by the al4α reading frame. In the present study we have used highly purified preparations of this enzyme to characterize its cleavage-recognition site, as well as its kinetic and substrate-binding properties, by using wild-type and mutant substrates derived from the al4α homing site of the *COX1* gene of *Saccharomyces capensis*. We also show that I-*Sce*II cleaves cloned human genomic DNA infrequently, suggesting that this enzyme may be suitable as a reagent for mapping studies of complex genomes.

* Corresponding author.

MATERIALS AND METHODS

Bacterial plasmids. The wild-type DNA substrate, pRSX, contains the I-*Sce*II recognition site consisting of 132 bp of aI3γ and 157 bp of fused exons 4 and 5 from the *COX1* gene of *S. capensis* (28, 29). An alternate wild-type substrate, pRSS, consists of a 29-bp synthetic oligonucleotide (which possesses the 18-bp recognition site) cloned into the *Sma*I-*Eco*RI sites of pUC18.

I-*Sce*II purification. I-*Sce*II was purified as previously described (29) except for one alteration in the preparation of fraction Ia. The enzyme supernatant fraction recovered after the addition of ammonium sulfate to 40% saturation was adjusted to 80% saturation by addition of solid ammonium sulfate. Stirring was continued for 30 min, and the precipitate, which contained the I-*Sce*II endonuclease activity, was collected by centrifugation and resuspended in buffer containing 10 mM potassium phosphate (pH 7.5), 45% glycerol, and 2 mM EDTA and stored at −20°C (fraction Ia).

I-*Sce*II endonuclease assay. Assays of I-*Sce*II activity were carried out by using plasmid DNA substrates linearized at a unique *Sca*I site and end labeled with [α-^{35}S]dATP as previously described (29). Reaction mixtures (25 μl) contained 25 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 50 mM KCl, 2 mM dithiothreitol, linearized plasmid DNA, and the enzyme sample. One unit of endonuclease activity is defined as the amount of enzyme that catalyzes the cleavage of 50 ng of the DNA substrate in 1 h at 30°C.

Human cosmid cloned plasmids. Human cosmids were isolated from several cosmid libraries constructed from somatic cell hybrids containing a single human chromosome 3 (8, 24, 25). Each cosmid sample was cleaved with *Eco*RI or *Pst*I and tested for the presence of an I-*Sce*II cleavage site by incubation with I-*Sce*II. Four cosmids (A-158, F-18, A-225, and F-95) were chosen for further characterization of their I-*Sce*II site. For each the I-*Sce*II site was mapped to a relatively small *Eco*RI or *Pst*I fragment, which was cloned into pBS(+). Further restriction digests and subcloning were conducted as needed to yield derivatives of pBS(+) in which the I-*Sce*II site from human DNA was situated near a universal primer site of the vector. These manipulations yielded I-*Sce*II sites in human DNA fragments of roughly 0.9 kb (A-158), 1.6 kb (F-18), 1 kb (A-225), and 1.3 kb (F-95).

Random mutagenesis of the I-*Sce*II recognition site. Random mutants with mutations spanning the I-*Sce*II recognition site were generated by the mutually primed oligonucleotide method of Hill et al. (10). The oligonucleotide 5′-CGGGATC CTTTGGTCATCCAGAAGTATACGAATTCG-3′ was synthesized on an Applied Biosystems oligonucleotide synthesizer such that the underlined sequences were added in a proportion of 90% of the designated base and 3.33% each of the other three nucleotides. The nonunderlined bases contained 100% of the designated nucleotide. The 3′ terminal 8 bases are palindromic and consequently can anneal to mutually prime DNA synthesis. The oligonucleotide (1 μg) was annealed at 70°C for 5 min and allowed to cool slowly to room temperature in a buffer containing 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol. DNA synthesis was initiated by the addition of deoxynucleoside triphosphates (dNTPs) to 250 μM and 10 U of the Klenow fragment of DNA polymerase I; the reaction was allowed to proceed at 23°C for 16 h. The double-stranded products were phenol extracted, ethanol precipitated, restricted with *Eco*RI, and cloned into the vector pUC18 digested with *Eco*RI and *Sma*I. Clones bearing an insert with one or only a few variant nucleotides were identified by colony hybridization with an oligonucleotide corresponding to the wild-type recognition sequence as a probe under relatively high-stringency conditions.

Site-directed mutagenesis of the I-*Sce*II recognition site. Site-directed mutants were generated by the method of Kunkel (12). Single-stranded DNA from plasmid pRSX was obtained by superinfecting the bacterial host CJS236 bearing this plasmid with the helper phage VCSM13 (Stratagene). Phosphorylated oligonucleotide (10 pmol) was annealed to 1 μg of template at 80°C in a buffer containing 50 mM NaCl in a volume of 20 μl; this was allowed to cool to room temperature slowly and was then shifted to ice. Dithiothreitol was added to 1 mM, dNTPs to 500 μM, and rATP to 500 μM. One unit of T4 DNA polymerase and 1 U of T4 DNA ligase were added. The reaction was allowed to proceed on ice for 5 min, at room temperature for 15 min, and at 37°C for 2 h. A 5-μl volume of this mixture were used to transform competent DH5α cells. Transformants were screened by sequencing by using Sequenase (United States Biochemical Corp.) and procedures recommended by the manufacturer.

Randomization of the I-*Sce*II cleavage site. The oligonucleotide 5′-ATGCGCAAGCTTGGTCNNNNAGAAGTATCGAT CCCTATAGTGAGTCGTATTA-3′ was synthesized on an Applied Biosystems oligonucleotide synthesizer to yield a sequence which contains the 18-base I-*Sce*II recognition site (boldface type) in which the nucleotides which make up the 4-base region known to encompass the I-*Sce*II cleavage site (N) are nearly completely randomized. This oligonucleotide also possesses the sequence (underlined) which is complementary to the T7 bacteriophage promoter. The synthetic oligonucleotide (0.4 μM) and the T7 promoter-primer (0.2 μM; Promega) were annealed by heating at 65°C for 10 min and cooling to room temperature. The primer-template (620 pmol as nucleotide) was extended by using 15 U of the *E. coli* DNA polymerase Klenow fragment (Promega) in the presence of all four dNTPs to produce a full-length double-stranded DNA fragment. The 52-bp fragment was isolated from a 2% agarose gel by electrophoresis onto an NA45 membrane (Schleicher & Schuell), eluted from the membrane, and digested with *Hin*dIII and *Cla*I to produce a 17-bp fragment with one 4-base (*Hin*dIII) and one 2-base (*Cla*I) 5′ overhanging end. The plasmid pGEM-4Z (Promega) was digested with *Acc*I and *Hin*dIII and ligated with the 21-nucleotide *Cla*I-*Hin*dIII fragment by incubation with T4 DNA ligase (Promega) at room temperature for 24 h. The ligation mixture was used to transform XL-1 cells (Stratagene), yielding 3015 transformants. The plasmid DNA was isolated from the pooled transformants and is designated pD7-10.

DNA sequencing. All of the plasmid DNAs generated in this study were sequenced by using the Sequenase Version 2.0 kit (United States Biochemical Corp.). The location of the cleavage site in the four human DNA clones noted above was determined by using these reagents and the approach described by Wenzlau et al. (28). The M13 primer was used for sequencing the site-directed mutants. Random mutants were sequenced by using a synthetic primer complementary to pUC18. The randomized I-*Sce*II site pGEM-4Z plasmid was sequenced with the SP6 promoter-primer (Promega) and an additional internal synthetic primer was used to sequence the second strand of the human clone, A-158.

PCR amplification of mutant and wild-type substrates. DNAs of selected mutant substrates for cleavage-competition experiments were prepared by polymerase chain reaction (PCR) amplification of 140 bp of pRSX and mutated versions of it by using two 21-base oligonucleotides, 5′-

TABLE 1. Comparison of I-*Sce*II cleavage-recognition sites

Site	Nucleotide at position[a]:																		% Identity[b]
	−1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
										↓										
S. capensis	T	T	T	G	G	T	C	A	T	C	C	A	G	A	A	G	T	A	T	100
S. cerevisiae	T	T	T	G	G	T	C	A	C	C	C	T	G	A	A	G	T	A	T	89
Random cleavage site (*S. capensis*)	C	T	T	G	G	T	C	N	N	N	N	A	G	A	A	G	T	A	T	
Cloned human sites																				
A-158	T	A	T	A	G	G	C	A	C	T	T	G	G	A	A	G	T	A	C	56
F-95	C	C	T	G	G	A	C	A	C	C	T	G	G	G	A	G	T	T	T	61
F-18	G	G	T	G	G	A	C	A	T	G	C	A	G	A	G	G	T	T	T	72
A-225	T	T	G	G	G	T	C	A	G	C	C	A	G	A	A	G	T	G	A	78
								↑												

[a] Shown are the I-*Sce*II sites in the *COX1* genes of *S. capensis* and *S. cerevisiae* and four human cloned DNAs. N indicates random nucleotides. The position of the 4-bp I-*Sce*II staggered cleavage is indicated by the arrows. The al4α intron insertion site is located between positions 5 and 6.
[b] Percent identity was calculated from positions 1 to 18 of the *S. capensis* site.

TGTAAGTGAAATATTTATATT-3′ and 5′-GAAATTTCA CCAAATACAGGT-3′, complementary to regions of the insert of pRSX (intron 3γ and exon 5, respectively) to generate a 140-bp fragment containing the I-*Sce*II recognition site in the middle. The oligonucleotides (0.35 μg each) and 0.2 μg of each linearized DNA substrate were incubated with the four dNTPs (200 μM each; Boehringer Mannheim) and 2.5 U of *Taq* DNA polymerase (Promega) in a Perkin-Elmer Cetus DNA Thermocycler. The DNA was phenol extracted, precipitated with ethanol, and resuspended in water, and the concentration was determined by measuring the UV absorbance.

RESULTS

A natural and synthetic I-*Sce*II site are cleaved with comparable kinetics. We have used highly purified preparations of active I-*Sce*II isolated from yeast mitochondria (29) to characterize interactions of the enzyme with wild-type and mutant substrates. This minimizes potential interference from contaminating proteins in crude mitochondrial extracts or unforeseen modifications of a mitochondrially encoded protein expressed from universal code equivalent genes in heterologous systems. Two substrates were used for the mutagenesis experiments described here, both derived from the *COX1* gene of *S. capensis*, a yeast which lacks al4α and whose mitochondrial genome is a natural recipient for that intron in crosses (28). The substrate pRSX, previously described by us (29), has a 289-bp fragment of the *COX1* gene containing the 18-bp I-*Sce*II recognition site, which spans the exon 4-5 junction; the other plasmid, pRSS, contains a synthetic 29-bp oligonucleotide containing the same 18-bp recognition site cloned into pUC18 (Table 1). This recognition site differs at positions 8 and 11 from the *S. cerevisiae* sequence used as the control in studies by Jacq and coworkers (5, 22) (Table 1). To confirm that the two substrates derived from *S. capensis* could be interchanged in the mutagenesis experiments, we compared the kinetics of their cleavage by I-*Sce*II by first linearizing each plasmid at the unique *Sca*I site and then incubating it with purified I-*Sce*II. Figure 1 shows that the initial velocity (Fig. 1A) and extent (Fig. 1B) of I-*Sce*II cleavage of the two substrates are comparable. Both have therefore been used in the mutagenesis experiments described here.

Most point mutations do not block cleavage by I-*Sce*II. Mutants with random and site-directed mutations of the I-*Sce*II recognition site were made as described in Materials and Methods; all point changes were verified by DNA sequencing. Eight point mutants were characterized from the set of random mutations of pRSS. Most of the mutants analyzed were obtained from site-directed mutations of pRSX. Altogether, 36 point mutants were analyzed to determine their effect on cleavage by purified preparations of I-*Sce*II. The results of these experiments (Fig. 2 and 3) show a range of cleavage efficiencies for the various mutant substrates, from those that are cleaved as well as the wild type is to those that are not cleaved at all. We have divided these mutants into four classes based on the efficiency of I-*Sce*II cleavage in a standard assay relative to the wild-type substrate. The kinetics of cleavage of an example of each class are shown in Fig. 2, and the relative cleavage efficiencies for all 36 mutants are summarized in Fig. 3. In all, 18 of the mutant substrates were cleaved at least 78% as well as the wild-type was (class I), whereas 6 others were cleaved at a somewhat reduced efficiency (42 to 60% of the wild-type level; class II). Nine mutations reduced the cleavage efficiency to 33% or less of the wild-type level (class III), and three of the mutations, G-4→C, G-12→T, and G-15→C, completely blocked cleavage under all conditions tested

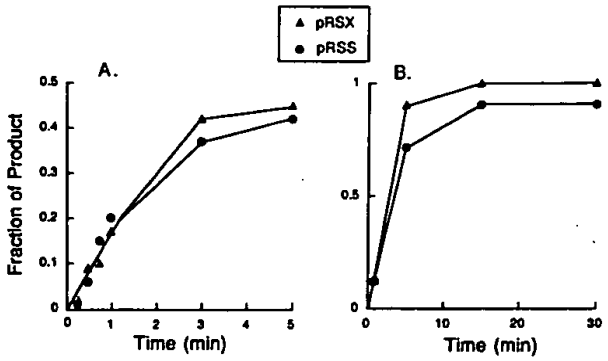


FIG. 1. The rate and extent of cleavage of the short (pRSS) and long (pRSX) forms of the wild-type DNA substrate (see Materials and Methods) by I-*Sce*II are nearly identical. (A) The linearized plasmids pRSX and pRSS (100 ng) were individually incubated with I-*Sce*II fraction Ia (11 U) for the indicated times. (B) The linearized plasmids pRSX and pRSS (0.25 ng) were individually incubated with I-*Sce*II fraction V (0.8 U) for the indicated times.

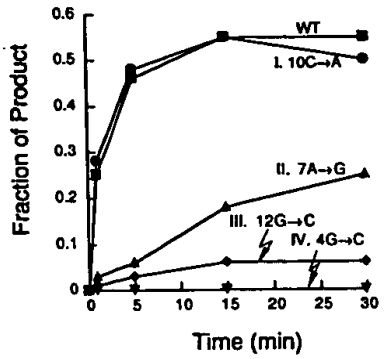


FIG. 2. I-*SceII* cleaves mutant DNA substrates with various degrees of efficiency. The linearized plasmids (100 ng) were incubated with I-*SceII* fraction V (12 U) for the indicated times. The data from 4 of the 36 individual mutant DNA substrates, representing each of the four classes of DNA substrate cleavage (I to IV), and the wild-type substrate, pRSX (WT), are shown.

(class IV). These data show the complexity of the I-*SceII* recognition site and, in particular, show that point mutations on both sides of the 4-bp cleavage site can profoundly affect the efficiency of cleavage by the enzyme.

Saturation mutagenesis of the cleavage site. In a mutational analysis of the substrate specificity of I-*SceI*, Colleaux et al. (4) found that each of six point mutations in the 4-bp I-*SceI* cleavage site blocked or greatly inhibited cleavage. Those results might suggest that I-*SceI* has a more stringent nucleotide requirement within its cleavage site than does I-*SceII*, since in our present data set (Fig. 3) only one of seven

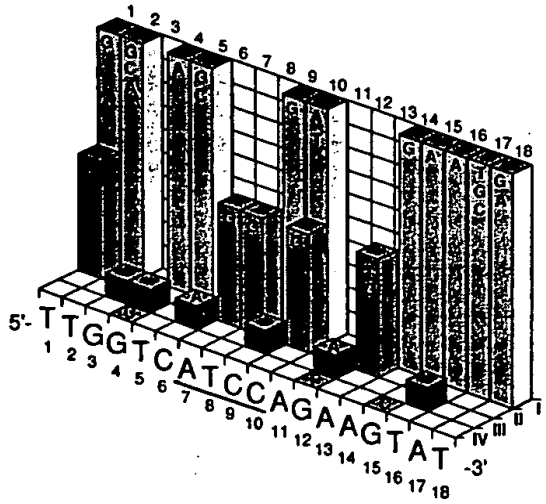


FIG. 3. Summary of the kinetics of mutant DNA substrate cleavage by highly purified I-*SceII* activity. All of the 36 mutant DNA substrates were assayed in the presence of I-*SceII* as described in Materials and Methods and the legend to Fig. 2. The results of these assays suggest four classes of cleavage efficiency as described in Results. The wild-type I-*SceII* recognition-cleavage site sequence and position number are indicated across the bottom of the figure. The four nucleotides which make up the 4-bp staggered cleavage site are underlined. Individual point mutant substrates tested and their relative efficiency of cleavage are indicated by the vertical bar and the relative height of the vertical bar, respectively.

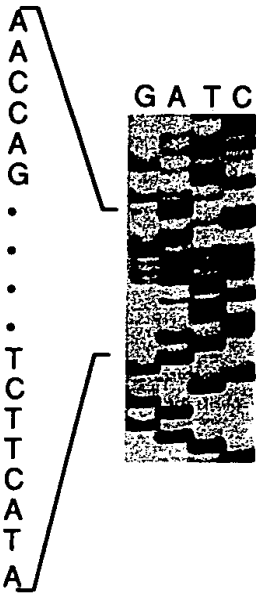


FIG. 4. The I-*SceII* cleavage site of pD7-10 is randomized. Construction of the I-*SceII* substrate, pD7-10, with a random nucleotide sequence at positions 7 to 10 is described in Materials and Methods. Sequencing indicates that the nucleotide composition of the degenerate I-*SceII* recognition site in pD7–10 is nearly random at positions 7 to 10.

mutations within the cleavage site (C-9→T) blocks cleavage substantially (to <5% of the wild-type substrate level). To obtain further insight into the nucleotide requirements within the 4-bp cleavage site of I-*SceII*, we synthesized an oligonucleotide that contains the 18-bp *S. capensis* recognition site but is degenerate at the cleavage site from positions 7 through 10 (Table 1) (see Materials and Methods). This generates a total of 256 possible combinations of the 4-bp cleavage site (including the *S. capensis* and *S. cerevisiae* wild-type sequences). DNA containing the degenerate I-*SceII* cleavage site was cloned into pGEM-4Z. More than 3,000 individual isolates from that cloning were pooled, and the DNA was extracted for sequence analysis. Figure 4 shows that the sequence of the 18-bp interval determined on the pooled DNA, pD7–10, is unambiguous except between positions 7 and 10, where all four nucleotides are present at each position. This result indicates that most, if not all, of the possible 256 nucleotide combinations within the 4-bp cleavage site are represented in the pooled DNA sample.

To determine the effects of the randomized cleavage site on I-*SceII* cleavage efficiency, we linearized 20-ng aliquots of the pooled plasmid DNAs containing the randomized cleavage site and the same amount of pRSX DNA containing the wild-type site with *ScaI* and incubated them with increasing amounts of I-*SceII*. Figure 5A shows that at an enzyme concentration sufficient to cleave 100% of the wild-type substrate, only about 10% of the plasmid containing the degenerate cleavage site was cut. This result suggests that possibly 25 sequences within the cleavage site permit cutting by I-*SceII*. As discussed below, these are likely to include many single and double mutations of the cleavage site. To rule out the possibility that at the concentrations of the various mutant DNAs present in the pooled sample, some are extremely effective as inhibitors of cleavage of an otherwise cleavable substrate, we have determined the ef-

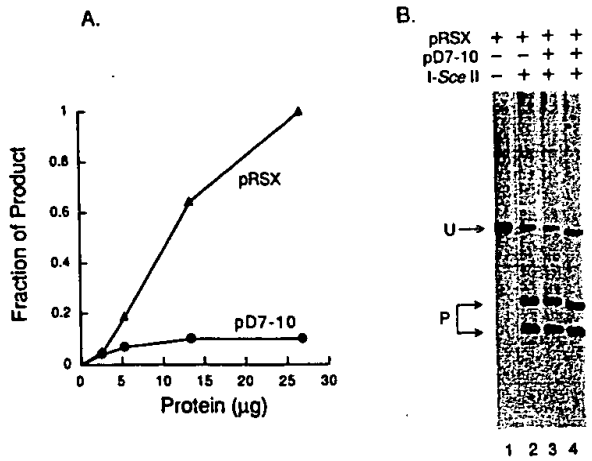


FIG. 5. (A) Kinetics of I-*Sce*II cleavage of the DNA substrate containing the degenerate I-*Sce*II cleavage site. The wild-type plasmid, pRSX (20 ng), and the degenerate plasmid, pD7–10 (20 ng), were linearized and incubated with I-*Sce*II fraction Ia (800 U/mg) as indicated. (B) Mixing of the wild-type plasmid, pRSX, and degenerate plasmid, pD7–10, DNA substrates. Radiolabeled pRSX DNA (20 ng) was assayed with I-*Sce*II fraction IIa (10 U) in the absence (lane 2) and presence (lanes 3 and 4) of unlabeled pD7-10 plasmid DNA. The number of potential cleavage sites contained within the pD7–10 DNA in lane 3 is 1.4 times that presented by the control pRSX substrate; that in lane 4 is 14 times the control number. The extents of pRSX DNA cleavage during reaction are as follows: lane 1 (no enzyme) 0; lane 2, 0.72; lane 3, 0.72; lane 4, 0.7. Abbreviations: U, uncut wild-type plasmid, pRSX; P, products of I-*Sce*II cleavage of pRSX.

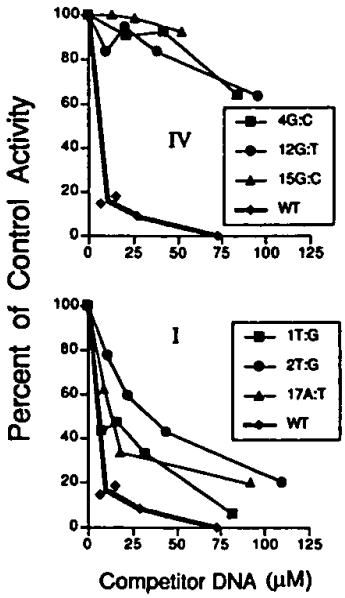


FIG. 6. Competition for I-*Sce*II cleavage by mutant substrates. The end-labeled wild-type substrate, pRSX (2.8 μM), was incubated with 0.6 U of I-*Sce*II (fraction V) and increasing concentrations of unlabeled competitor DNAs, which were generated by PCR as described in Materials and Methods.

fects of the pooled, degenerate cleavage site DNAs on I-*Sce*II cleavage of the wild-type substrate. In this experiment, unlabeled pD7-10 DNA with the degenerate cleavage site were added to end-labeled pRSX and the mixture was incubated with an amount of enzyme that would yield about 75% cleavage of pRSX alone (Fig. 5B, lane 2). The results of this experiment show that addition of a 1.4- and 14-fold excess of the unlabeled DNA had no effect on the cleavage of pRSX (lanes 3 and 4), indicating that the inefficient cleavage of the substrate with the degenerate cleavage site is not due to some inhibitor present in that material. Although we have not carried out an extensive analysis of specific sequences within the cleavage site that either permit or block I-*Sce*II cutting in this experiment, we have retrieved the wild-type *S. capensis* cleavage site from the pooled DNAs and found that it is cleaved indistinguishably from the 289-bp form of the wild-type substrate in pRSX (data not shown). These experiments demonstrate, therefore, that the large majority of permutations of the cleavage site severely affect the ability of I-*Sce*II to cleave there.

Competition for I-*Sce*II cleavage by mutant substrates. The above experiments show that mutations both within and outside of the cleavage site have a wide range of effects on the efficiency of I-*Sce*II cleavage. To determine whether these effects correlate with the ability of the enzyme to bind to the mutant substrate, we have carried out competition experiments between selected mutants and the wild-type substrate. To be able to use a wide concentration range of the various mutant substrates in these experiments, we used PCR to amplify their recognition sites. A total of six mutant DNAs (three each of classes I and IV; Fig. 3) and the wild-type *S. capensis* substrates were amplified with two PCR primers designed to generate 140-bp fragments containing the 18-bp recognition site (see Materials and Methods). These amplified DNAs were incubated at various concentrations with a constant amount of ^{35}S-end-labeled pRSX DNA, which was present at saturating amounts for I-*Sce*II cleavage; in this way competition by unlabeled substrate would be evident as a reduction in the amount of labeled product produced under the standard reaction conditions.

The results of these experiments (Fig. 6) show a strong correlation between the ability of a mutant substrate to be cleaved by I-*Sce*II and its ability to compete for cleavage of the wild-type substrate. The most effective competitor is the PCR-amplified wild-type substrate. Of the three class I mutants that are cleaved indistinguishably from the wild type, two (T-1→G and A-17→T) are good competitors whereas one (T-2→G) is somewhat less effective as an inhibitor. These data suggest that each mutant substrate may interact differently with the enzyme, even in cases when cleavage appears unaltered. All three of the class IV mutants which fail to be cleaved by I-*Sce*II, however, show no appreciable inhibition of cleavage of the wild-type substrate until very high concentrations of competitor DNA are used (an effect most probably due to nonspecific inhibition of activity). Although these experiments do not rule out the possibility of significant nonproductive binding of I-*Sce*II to some noncleavable or poorly cleavable mutant form of the substrate, they do show that the point changes in the recognition site can alter I-*Sce*II binding and hence its ability to cleave the substrate.

I-*Sce*II cleaves cloned human DNA infrequently. Thierry et al. (27) have recently shown that I-*Sce*I does not cleave yeast nuclear DNA (1.4×10^7 bp), consistent with their earlier conclusion of the high degree of complexity (and specificity) of its cleavage-recognition site (4). Our analysis of the cleavage-recognition site of I-*Sce*II, together with

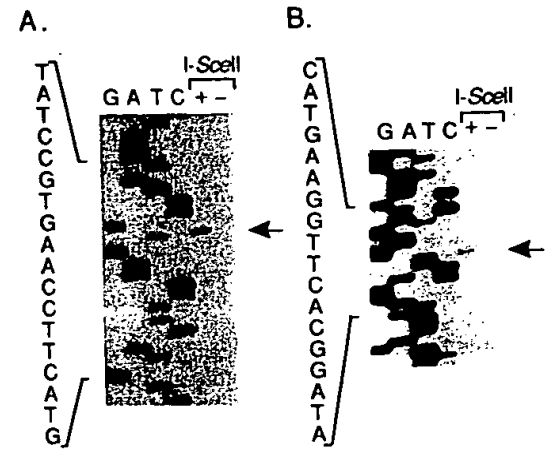


FIG. 7. Sequence of the I-*Sce*II cleavage site in the human chromosome 3 cosmid clone, A-158. Sequence analysis and mapping of the I-*Sce*II cleavage site were determined as described by Wenzlau et al. (28). Complementary strands are shown in panels A and B. Arrows indicate sites of I-*Sce*II cleavage.

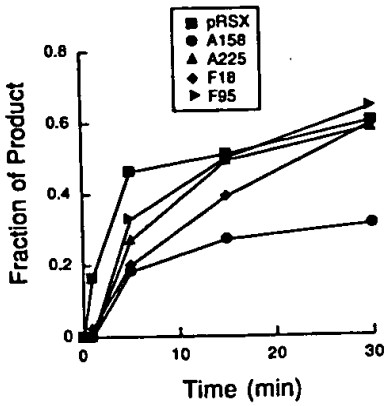


FIG. 8. Kinetics of I-*Sce*II cleavage of cloned human chromosome 3 DNA. Radiolabeled DNA from each of the four human chromosome 3 cosmid clones (100 ng) and both of the wild-type substrates (100 ng) was individually incubated with I-*Sce*II fraction III (12 U) as indicated.

findings of Sargueil et al. (22), suggest that I-*Sce*II should cleave complex genomes more frequently than does I-*Sce*I.

To assess the utility of I-*Sce*II as a reagent for physical mapping of complex genomes, we have used various preparations of I-*Sce*II activity from yeast mitochondria to screen cosmid-cloned DNAs containing fragments of human chromosome 3 for I-*Sce*II cleavage sites. Of 207 cosmids screened, representing more than 7 Mbp of human DNA, 144 had no apparent I-*Sce*II sites, 62 had one site, and 1 had two sites. Thus, there is one I-*Sce*II site for every 110 kbp in this sample of human DNA, a much higher frequency than would be expected if there were a stringent requirement for the 18-bp wild-type *S. capensis* or *S. cerevisiae* I-*Sce*II recognition site. These results are consistent with our point mutant analysis in which many point mutations did not block cleavage.

To confirm that the above inferences are correct, we characterized the I-*Sce*II cleavage site in four of the cosmids. First, in each cosmid the I-*Sce*II site was mapped relative to cleavage sites of standard restriction enzymes, and then a small fragment from each containing the I-*Sce*II site was subcloned into pBS(+). The cleavage site of each of those clones was mapped by using the method devised by us previously (28). Figure 7 shows the primary sequence of both strands of the subclone from cosmid A-158 surrounding the I-*Sce*II site and the site of cleavage. From those data the sequence surrounding the cleavage is established as summarized in Table 1: clearly, like its action on yeast mitochondrial substrates, I-*Sce*II makes a 4-base staggered cut, leaving 3' overhangs. The sequence around the A-158 site matches the *S. capensis* site at 10 of 18 positions. The three other sites, from cosmids F-95, F-18, and A-225, were mapped on one strand in this way. In all four clones, the mapped cleavage site occurs at the same position relative to the yeast cleavage site (Table 1). These data clearly show that human DNA contains sites for I-*Sce*II cleavage and that those sites resemble each other and the known sites from yeast mitochondrial DNA. Using plasmids containing these fragments of human DNA, we compared the kinetics of cleavage of each of these DNAs (Fig. 8). Each is a good substrate, being cleaved at no less than 50% the rate of the control substrate, despite the presence of four to eight point mutations relative to the yeast sites.

DISCUSSION

Using highly purified preparations of I-*Sce*II endonuclease isolated from an *S. cerevisiae* strain that overproduces the enzyme, we have analyzed the efficiency of cleavage of a set of point mutations within the I-*Sce*II recognition site, previously defined as 18 bp spanning the junction of exons 4 and 5 of the mitochondrial *COX1* gene of *S. capensis* (28). Of 36 point mutations within those 18 bp, we find that 18, including 3 within the 4-bp cleavage site, have little or no effect on cleavage efficiency. Six of the point mutations reduce the efficiency of I-*Sce*II cleavage to 42 to 60% of that of the wild type, and nine reduce cleavage to less than 33% of control levels. Only three mutations, at positions 4, 12, and 15, all of which are outside of the 4-bp cleavage site, block cleavage completely. The results of competition experiments suggest that, not surprisingly, point mutations which severely reduce I-*Sce*II cleavage can do so by reducing the affinity of the enzyme for the substrate DNA. Since we have not exhaustively surveyed the ability of the various mutant substrates to compete with the wild-type substrate for cleavage by I-*Sce*II, we cannot rule out the possibility that some mutations affect primarily k_{cat} rather than K_m.

Sargueil et al. (22) have carried out a similar analysis of the effects of substrate point mutations on I-*Sce*II cleavage. The substrates used in their study were derived from the *COX1* gene of *S. cerevisiae*, which differs from the *S. capensis* gene at two positions (using the numbering system described in the present work): one within the cleavage site at +8 (C→T) and the other just outside of the cleavage site at position +11 (T→A) (Table 1). The I-*Sce*II activities used in that study were preparations obtained from extracts of *E. coli* harboring universal code-equivalent genes of I-*Sce*II and from crude extracts of mitochondria of wild-type yeast cells. For unknown reasons, the enzyme activities expressed from universal code-equivalent genes in *E. coli* showed a very different and significantly more relaxed cleavage specificity than did the enzyme activity they extracted from yeast mitochondria (22). Therefore, we will restrict comparisons of their point mutation data set with ours to assays done with

I-*Sce*II activity extracted from yeast mitochondria. Comparison of identical point mutations within the 18-bp recognition site (a total of 22) shows that the two data sets are in reasonable agreement, with the exception of G-4→A. In that case, Sargueil et al. (22) reported that the mutation in the *S. cerevisiae* substrate completely blocked cleavage by I-*Sce*II whereas the same change in the *S. capensis* substrate site as shown here has little effect on cleavage.

There is one other significant difference between the two studies. Sargueil et al. (22) reported that I-*Sce*II cleavage is blocked completely when the T at the −1 position of the *S. cerevisiae* substrate (also present in the long *S. capensis* sequence) is replaced with a G or C. However, a substrate containing the *S. capensis* I-*Sce*II recognition site retrieved from the random mutagenesis of the cleavage site contains a C at the −1 position and is cleaved by I-*Sce*II with the same kinetics as the natural substrate with a T at −1. This difference is particularly intriguing since our initial delineation of the I-*Sce*II recognition site in *S. capensis* was based on a determination of the minimum length of double-stranded DNA around the intron insertion site necessary for I-*Sce*II cleavage and did not include the nucleotide pair at the −1 position.

Further insight into the complexity of the I-*Sce*II site comes from the analysis of I-*Sce*II cleavage sites in human DNA. All four of the human sequences are distinct, but each clearly resembles the yeast mitochondrial site; for example, the best match among the human sites is 14 of 18 identical positions and the worst match is 10 of 18 positions. Six positions are invariant among the four new sites. The four human sites vary more among themselves than they do relative to the yeast site; for example, the most similar pair, A-158/F-95 and F-95/F-18, are identical at only 11 of 18 positions, whereas the least similar pair, A-158/F-18, is identical at only 7 of 18 positions. Significantly, the four human cosmid clones we have analyzed are cleaved efficiently by I-*Sce*II even though two of them, F-18 and F-95, have a G and a C, respectively, at the −1 position (Table 1). Moreover, at position 9, clone A-158 has a T residue which nearly blocks cleavage of the *S. capensis* substrate. It also should be noted that the three sites at which point mutations block cleavage of the *S. capensis* substrate (positions 4, 12, and 15) are invariant among the human sites. Position 6 is also invariant, and it may be relevant that all three point mutations there block cleavage of the *S. capensis* substrate nearly completely. Although these observations provide some additional clues to I-*Sce*II sequence specificity, we cannot conclude from this small data set that invariant positions are absolute predictors of I-*Sce*II cleavage specificity. At position 16, for example, all of the human sites have a T residue, like the control substrate, even though our point mutation data indicate that an A residue at that position has no effect on cleavage.

At present there is only a limited data base on point-mutational analyses of double-stranded DNA endonuclease substrates with long recognition sites comparable to I-*Sce*II. In one such study of I-*Sce*I, most point mutations of the recognition site for that intron-encoded endonuclease block or significantly reduce cleavage by the enzyme and five of six within the 4-bp cleavage site were reported to block cleavage completely (4). Our data, together with those of Sargueil et al. (22), indicate that, unlike the I-*Sce*I substrate, most of the point mutations within the I-*Sce*II substrate recognition site do not severely compromise I-*Sce*II cutting efficiency. For example, 8 of a total of 12 possible mutants with point mutations of the *S. capensis* cleavage site that have been analyzed (including the *S. cerevisiae* site) are reasonably good substrates. Considering these data together with the results of the analysis of the cloned human DNA, it is likely that a significant number of the cleavable mutant substrates derived from the random mutagenesis of the cleavage site are single and double mutations. In all, it appears that I-*Sce*II has a significantly less stringent sequence requirement for cleavage than does I-*Sce*I, a conclusion also suggested by the observation that in 1.4×10^7 bp of DNA (representing the yeast nuclear genome) there is no I-*Sce*I cleavage site (27), whereas we find that DNA from a chromosome of another complex genome (human) contains an I-*Sce*II site every 1.1×10^5 bp.

One of the interesting aspects of this class of site-specific endonucleases with complex recognition sites is the distinction between the sequences in and around the cleavage site and the overall recognition site for the enzyme. For example, the results of mutational analysis of the HO endonuclease recognition site in the *MAT* locus of yeast nuclear DNA shows that the integrity of a core of ca. 8 bp spanning the 4-bp cleavage site (in the z region of the *MAT* site) is required for HO endonuclease cutting (19). Most point changes outside of that core have little effect on cleavage, whereas multiple changes outside of the core severely affect the efficiency of cleavage by HO endonuclease.

A comparison of the point mutant data for the *S. cerevisiae* and *S. capensis* I-*Sce*II sites and the sites in human DNA raises the interesting possibility that for any given mutation in the I-*Sce*II cleavage-recognition site, the sequence context could dramatically affect the efficiency by which a given mutant substrate is cleaved by the enzyme. This could mean that I-*Sce*II recognizes or is sensitive to some conformation of its substrate, which could well depend on the juxtaposition of multiple nucleotides that make up a functional recognition-cleavage site, as appears to be the case for the HO endonuclease site (19). Experiments to examine these possibilities are in progress.

ACKNOWLEDGMENTS

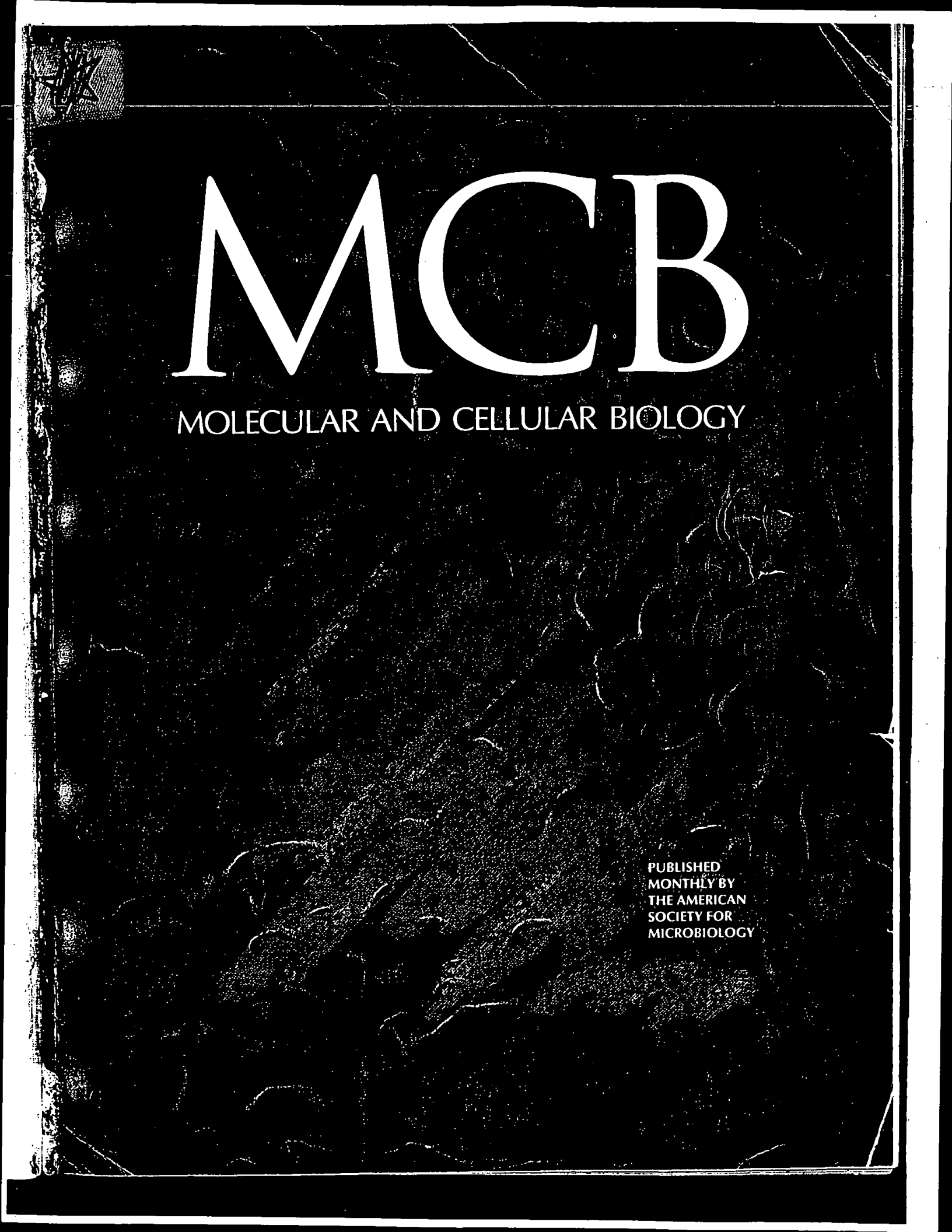
We thank Janet Wenzlau for sequencing some of the original point mutants of the I-*Sce*II recognition site. We also thank Betty Key for expert technical assistance.

This work was supported by NIH grant GM35510 to R.A.B. and P.S.P. and by grant I-642 from the Robert A. Welch Foundation and NIH grant CA48031 to D.I.S. Catherine Wernette was supported by a postdoctoral fellowship from the NIH.

REFERENCES

1. **Bell-Pederson, D., S. Quirk, M. Aubrey, and M. Belfort.** 1989. A site-specific endonuclease and co-conversion of flanking exons associated with the mobile *td* intron of phage T4. Gene 82:119–126.
2. **Bell-Pedersen, D., S. Quirk, J. Clyman, and M. Belfort.** 1990. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res. 18:3763–3770.
3. **Colleaux, L., L. d'Auriol, M. Betermier, G. Cottarel, A. Jacquier, F. Gailbert, and B. Dujon.** 1986. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease. Cell 44:521–533.
4. **Colleaux, L., L. d'Auriol, F. Galibert, and B. Dujon.** 1988. Recognition and cleavage site of the intron-encoded omega transposase. Proc. Natl. Acad. Sci. USA 85:6022–6026.
5. **Delahodde, A., V. Goguel, A. M. Becam, F. Creusot, J. Banroques, and C. Jacq.** 1989. Site specific DNA endonuclease and RNA maturase activities of two homologous intron encoded

proteins from yeast mitochondria. Cell 56:431–441.
6. **Dujon, B., M. Belfort, R. A. Butow, C. Jacq, C. Lemieux, P. S. Perlman, and V. M. Vogt.** 1989. Mobile introns: definition of terms and recommended nomenclature. Gene 82:115–118.
7. **Gauthier, A., M. Turmel, and C. Lemieux.** 1991. A group I intron in the chloroplast large subunit rRNA gene of *Chlamydomonas eugametos* encodes a double-strand endonuclease that cleaves the homing site of this intron. Curr. Genet. 19:43–47.
8. **Golembieski, W. A., S. E. Smith, F. Recchia, A. Judge, V. Shridhar, O. J. Miller, H. Drabkin, and D. I. Smith.** 1991. Isolation of large numbers of chromosome 3-specific cosmids containing clusters of rare restriction endonuclease sites. Am. J. Hum. Genet. 49:581–589.
9. **Hanson, D. K., M. R. Lamb, H. R. Mahler, and P. S. Perlman.** 1982. Evidence for translated intervening sequences in the mitochondrial genome of Saccharomyces cerevisiae. J. Biol. Chem. 257:3218–3224.
10. **Hill, D. E., A. R. Oliphant, and K. Struhl.** 1982. Mutagenesis with degenerate oligonucleotides: an efficient method for saturating a defined DNA region with base pair substitutions. Methods Enzymol. 155:558–568.
11. **Jacquier, A., and B. Dujon.** 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41:383–394.
12. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA 82:488–492.
13. **Lemieux, B., M. Turmel, and C. Lemieux.** 1989. Unidirectional gene conversions in the chloroplast of Chlamydomonas interspecific hybrids. Mol. Gen. Genet. 212:48–55.
14. **Lemieux, C., and R. W. Lee.** 1987. Nonreciprocal recombination between alleles of the chloroplast 23S rRNA gene in interspecific Chlamydomonas crosses. Proc. Natl. Acad. Sci. USA 84:4166–4170.
15. **Macreadie, I. G., R. M. Scott, A. R. Zinn, and R. A. Butow.** 1985. Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. Cell 41:395–402.
16. **Monteilhet, M., A. Perrin, A. Thierry, L. Colleaux, and B. Dujon.** 1990. Purification and characterization of the *in vitro* activity of I-*Sce*I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acids Res. 18:1407–1413.
17. **Muscarella, D. E., E. L. Ellison, B. M. Ruoff, and V. M. Vogt.** 1990. Characterization of I-PPO, an intron-encoded endonuclease that mediates homing of a group-I intron in the ribosomal DNA of *Physarum polycephalum*. Mol. Cell. Biol. 10:3386–3396.
18. **Muscarella, D. E., and V. M. Vogt.** 1989. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. Cell 56:443–454.
19. **Nickoloff, J. A., J. D. Singer, and F. Heffron.** 1990. In vivo analysis of the *Saccharomyces cerevisiae* HO nuclease recognition site by site-directed mutagenesis. Mol. Cell. Biol. 10:1174–1179.
20. **Perlman, P. S., and R. A. Butow.** 1989. Mobile introns and intron-encoded proteins. Science 246:1106–1109.
21. **Quirk, S. M., D. Bell-Pedersen, and M. Belfort.** 1989. Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. Cell 56:455–465.
22. **Sargueil, B., D. Hatat, A. Delahodde, and C. Jacq.** 1990. *In vivo* and *in vitro* analyses of an intron-encoded DNA endonuclease from yeast mitochondria. Recognition site by site-directed mutagenesis. Nucleic Acids Res. 18:5659–5665.
23. **Sass, P. M.** 1984. Mutations in the mosaic *oxi1* gene in the mitochondrial genome of the yeast *Saccharomyces cerevisiae*. Ph.D. thesis, Indiana University, Bloomington.
24. **Smith, D. I., W. Liu, D. Ginzinger, P. Green, S. Smith, N. Wang, F. Recchia, K. Carolyn, H. A. Drabkin, and W. Golembieske.** 1991. Localization of 616 human chromosome 3-specific cosmids using a somatic cell hybrid deletion mapping panel. Genomics 11:179–187.
25. **Smith, D. I., R. Mangrulker, R. Deisi, J. Gilbert, K. Kinsman, H. Drabkin, and W. Golembieske.** 1989. Saturation of human chromosome 3 with unique sequence hybridization probes. Genomics 4:453–459.
26. **Szostak, J., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl.** 1983. The double strand break repair model for recombination. Cell 33:25–35.
27. **Thierry, A., A. Perrin, J. Boyer, C. Fairhead, B. Dujon, B. Frey, and G. Schmitz.** 1991. Cleavage of yeast and bacteriophage-T7 genomes at a single site using the rare cutter endonuclease I-*Sce*I. Nucleic Acids Res. 19:189–190.
28. **Wenzlau, J. M., R. J. Saldanha, R. A. Butow, and P. S. Perlman.** 1989. A latent intron-encoded maturase is also an endonuclease needed for intron mobility. Cell 56:421–430.
29. **Wernette, C. M., R. Saldanha, P. S. Perlman, and R. A. Butow.** 1990. Purification of a site-specific endonuclease, I-*Sce*II, encoded by intron 4α of the mitochondrial *cox1* gene of *Saccharomyces cerevisiae*. J. Biol. Chem. 265:18976–18982.
30. **Zinn, A. R., and R. A. Butow.** 1985. Nonreciprocal exchange between alleles of the yeast mitochondrial 21S rRNA gene: kinetics and the involvement of a double-strand break. Cell 40:887–895.


MCB
MOLECULAR AND CELLULAR BIOLOGY
PUBLISHED
MONTHLY BY
THE AMERICAN
SOCIETY FOR
MICROBIOLOGY

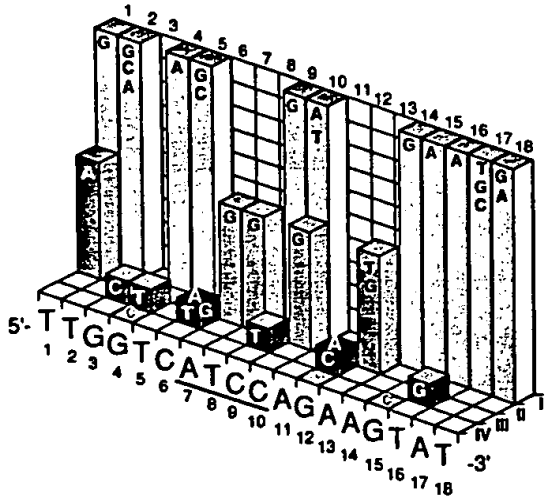
ERRATUM

Complex Recognition Site for the Group I Intron-Encoded Endonuclease I-*Sce*II

CATHERINE WERNETTE, ROLAND SALDANHA, DAVID SMITH, DING MING, PHILIP S. PERLMAN, AND RONALD A. BUTOW

Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235; Department of Molecular Genetics, The Ohio State University, Columbus, Ohio 43210; and Department of Molecular Biology and Genetics, Wayne State University School of Medicine, Detroit, Michigan 48201

Volume 12, no. 2, p. 719: Figure 3 should appear as shown below. The white letters did not show clearly in the original article.



MCB

APR. 1992
VOLUME 12
NUMBER 4

MOLECULAR AND CELLULAR BIOLOGY

PUBLISHED MONTHLY BY THE AMERICAN SOCIETY FOR MICROBIOLOGY